DOCUMENT RESUME

ED 230 183 IR 010 704

AUTHOR Lockhart, Kathleen A.; And Others
TITLE The Effects of the Timing of Feedback on Long-Term Knowledge Retention in PSI Courses. Final Report.
INSTITUTION Navy Personnel Research and Development Center, San Diego, Calif.
REPORT NO NPRDC-TR-83-13
PUB DATE Feb 83
NOTE 34p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cost Effectiveness; *Feedback; Higher Education; *Long Term Memory; Retention (Psychology); *Test Format; *Test Items; Test Results; Test Use; Undergraduate Students
IDENTIFIERS *Personalized System of Instruction

ABSTRACT

Three experiments were conducted, all employing undergraduates in college courses taught according to personalized system of instruction (PSI) principles. Experiment I examined retention as a function of the feedback delay interval in an introductory anthropology course using short-answer essay tests. Experiment II varied the feedback delay interval, the informational quality of the feedback, and test item type, and Experiment III examined delay and item type in a psychology course on experimental design. Results showed that neither delayed nor immediate feedback produced superior, long-term knowledge retention, regardless of feedback quality. Further, subjects in the immediate feedback conditions did not repeat initial errors more frequently than did delay subjects. The frequent, repeatable-quizzing aspect of PSI may make feedback a less potent variable than in other courses, and results do not support the use of immediate feedback when cost considerations and convenience or administration are important. Fourteen references and 16 data tables are included. (LMM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☒ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

NPRDC TR 83-13

February 1983

ED230183

THE EFFECTS OF THE TIMING OF FEEDBACK ON LONG-TERM KNOWLEDGE RETENTION IN PSI COURSES

Kathleen A. Lockhart
Persis T. Sturges
Judith Zachai
Barbara Dubois
Dawn Groves

Reviewed by
J. C. McLachlan

Released by
James F. Kelly, Jr.
Commanding Officer

Navy Personnel Research and Development Center
San Diego, California 92152

IR010704

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPRDC TR 83-13	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE EFFECTS OF THE TIMING OF FEEDBACK ON LONG-TERM KNOWLEDGE RETENTION IN PSI COURSES		5. TYPE OF REPORT & PERIOD COVERED Final Report Oct 1979-Sep 1981 6. PERFORMING ORG. REPORT NUMBER 14-81-25
7. AUTHOR(s) Kathleen A. Lockhart, Barbara Dubois, Persis T. Sturges, Dawn Groves, Judith Zachai		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ZR000-01-042-06.01.02
11. CONTROLLING OFFICE NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		12. REPORT DATE February 1983 13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Knowledge retention
Personalized system of instruction
Delayed feedback
Computer-managed instruction
Programmed instruction

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Several widely used instructional delivery systems include the provision of immediate feedback of test results as a crucial component. Given the cost of such feedback, the instructional benefits obtained should be verified. This report describes a series of three experiments concerning the effects of timing of feedback on long-term knowledge retention in courses taught according to the personalized system of instruction.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Experiment I varied only the timing of the feedback. Experiment II varied the timing of feedback, the quality of feedback, and test-item type. Experiment III varied the timing of feedback and test-item type. Analysis of the data showed that neither delayed or immediate feedback produced superior long-term knowledge retention, regardless of the quality of feedback (within the limitation of these experiments). These results do not support the use of immediate feedback when cost considerations are important. Delayed feedback was as effective in this series of experiments, and is far easier to implement in operational training.

FOREWORD

This research was conducted within the Interlaboratory Independent Research Program under work unit ZR000-01-042-06.01.02 (Delayed Feedback in Acquisition and Retention). This report describes the results of a series of three experiments examining the relationship between the timing of feedback and long-term knowledge retention. It is intended primarily for researchers working in the area of delayed feedback. However, the results and conclusions should be of interest to those concerned with designing instructional delivery systems, including computer-managed instruction, programmed instruction, and the personalized system of instruction.

JAMES F. KELLY, JR.
Commanding Officer

JAMES W. TWEEDDALE
Technical Director

CONTENTS

	Page
INTRODUCTION	1
Problem	1
Objective	1
Background	1
Experimental Conditions	3
EXPERIMENT I	3
Approach	3
Experimental Design and Subjects	3
Test Schedule and Materials	3
Experimental Feedback	4
Study Time Sheets	4
Proctors	4
Procedure	4
Analyses	5
Results of Experiment I	5
Reliability of Scoring the Short Essay Answers	5
Group Performance on Initial Attempts on Quizzes	5
Group Performance on Review Tests	6
Group Performance on Final Exam	6
Study Time	6
Number of Remediations	6
Proportions of Items that were Incorrect or Correct on a Review Test that were Correct on the Final Exam	7
EXPERIMENT II	8
Approach	8
Experimental Design	8
Subjects	8
Test Schedule	9
Test Materials	9
Criterion	10
Objectives	10
Tutoring	10
Procedure	11
Analyses	12
Results of Experiment II	12
Reliability of Scoring Fill-in Test Items	12
Group Performance on Study Quiz 1	12
Group Performance on Multiple-choice Items	13
Group Performance on Fill-in Items	14
Proportions of Items that were Incorrect or Correct on a Study Quiz that were Correct on the Next Review Test	15
Proportions of Items that were Incorrect or Correct on a Review Test that were Correct on the Final Exam	16

Page

EXPERIMENT II 17

Approach 17
Experimental Design and Subjects 17
Test Schedule 18
Testing Materials 18
Tutoring 19
Procedure 19
Analyses 20
Results of Experiment III 20
Group Performance on Study Quiz 1 20
Group Performance on Multiple-choice Items on Study Quizzes and Review Test 20
Group Performance on Fill-in Items on Study Quizzes and Review Tests 21
Loss from Review Test to Final Exam for Multiple-choice and Fill-in Items 21
Number of Remediations 21
Proportion of Items that were Incorrect or Correct on a Study Quiz that were Correct on the Next Review Test 22
Proportion of Items that were Incorrect or Correct on Review Tests that were Correct on the Final Exam 23

CONCLUSIONS 24

RECOMMENDATION 25

REFERENCES 27

DISTRIBUTION LIST 29

LIST OF TABLES

		Page
1.	Experimental Quiz/Test Schedule (Experiment I)	4
2.	Mean Correct on Unit Quizzes, Review Tests, and Final Exam (Experiment I)	6
3.	Group Means for Total Study Time and Number of Remediations (Experiment I)	7
4.	Proportion of Items Correct and Incorrect on the Review Tests that were Correct on the Final Exam (Experiment II)	7
5.	Schedule (Experiment II)	9
6.	Mean Number of Correct Answers on Study Quiz I (Experiment II)	12
7.	Mean Number Correct for Items on Review tests and Final Exam (Experiment II)	13
8.	Proportions of Multiple-choice Items Incorrect or Correct on a Study Quiz that Were Correct on the Next Review Test (Experiment II)	15
9.	Proportions of Fill-in Items Incorrect or Correct on a Study Quiz that Were Correct on the Next Review Test (Experiment II)	16
10.	Proportions of Items Incorrect or Correct on a Review Test that Were Correct on the Final Exam (Experiment II)	17
11.	Schedule (Experiment III)	18
12.	Mean Numbers Correct for Items on Review Tests and Final Exam (Experiment III)	20
13.	Mean Numbers of Remediations (Experiment III)	22
14.	Proportion of Multiple-choice Items that Were Incorrect or Correct on the Study Quizzes that Were Correct on the Next Review Test (Experiment III)	22
15.	Proportions of Fill-in Items that Were Incorrect or Correct on the Next Review Test (Experiment III)	23
16.	Proportions of Items Incorrect or Correct on Review Test that Were Correct on the Final Exam (Experiment III)	23

INTRODUCTION

Problem

The personalized system of instruction (PSI), precision teaching, and the Navy's computer-managed instruction (CMI) system are among several instructional systems that dictate the provision of immediate feedback in order to maximize student learning. However, immediate feedback is expensive in student time and in instructor or proctor time and evidence is mounting that delayed feedback produces equal learning and frequently superior retention, at least when multiple-choice or fill-in test items are used. The issue remains as to what the optimal feedback procedures are for these kinds of instructional systems.

Objective

The objective of this series of experiments was to examine the relationship between the timing of feedback and long-term knowledge retention under classroom conditions that exist in courses taught according to the principles of PSI.

Background

Many of the recent innovations in instruction have provided for immediate feedback of test results to the students. The feedback typically included information concerning the accuracy of answers and it may also contain additional material designed to allow the students to correct their errors. In the Navy's CMI system (Van Matre, 1980), for example, tests are scored by the computer upon test completion, and feedback consists of an indication of the correctness of the answer, as well as materials that the students should consult to correct their mistakes. Keller's (1968) PSI uses proctors to provide immediate feedback that consists of an indication of the correctness of an answer and remedial assignments to help students find the answers to items they missed.

Obviously, these systems devote considerable effort and expense to ensuring that students receive knowledge of results immediately. This is true despite the fact that there is considerable evidence demonstrating the superior efficacy of delayed feedback, at least in terms of long-term knowledge retention. The question remains as to what the optimal feedback procedures are for PSI type courses.

Before discussing the existing evidence regarding feedback effectiveness, it is necessary to review recent research in this area. The typical experiment has used two groups of subjects. After initial exposure to the test material in the form of multiple-choice questions, feedback of results has been provided either immediately, or following a delay of some interval. After a retention interval, both groups of subjects received the same tests again. Using this basic design, Sassenrath and Yonge (1968, 1969), Sturges (1969, 1972, 1978), and Kulhavy and Anderson (1972) all demonstrated that delayed feedback produced superior retention when compared to immediate feedback. These same studies also showed that there was no difference in immediate acquisition as a function of feedback delay interval. Further, the validity of the phenomenon has been studied in several experiments employing students in classroom settings and procedures such as would be found in a standard educational environment. Moore (1969), Sturges (1972), and Surber and Anderson (1975) all demonstrated the superiority of delayed feedback in classroom settings.

Several explanations have been tendered to account for this "delay-retention effect." Sturges (1972) suggests that subjects receiving delayed feedback either learn to discriminate the correct choice more precisely (because they learn both the correct and the incorrect alternatives from the feedback) or they engage in higher order organization of the information. Her data support the latter interpretation. It appears that subjects in immediate feedback conditions examine feedback only sufficiently to determine whether their answers are right or wrong. Delayed feedback subjects, however, usually must study all the feedback to remember the question and their answer. In either case, Sturges hypothesizes that the crucial period is the period after the subject receives the feedback, not the delay interval per se.

Kulhavy and Anderson (1972) hypothesize that proactive interference accounts for the differences in retention. Subjects in the delay condition forget their errors so that they are able to learn the correct answers when they receive feedback. Subjects in the immediate feedback conditions are perseverating on their incorrect answers; therefore, interference prevents them from acquiring the correct response. Support for this hypothesis is evidenced in the Kulhavy and Anderson experiments that show that the probability of repeating initial errors on the retention test is greater for subjects in the immediate feedback condition than for those in the delay condition.

Few people currently suggest that reinforcement theory adequately accounts for the effects of feedback. Keller's PSI approach was, of course, an attempt to implement the principles of operant conditioning in the classroom. In the effort to accomplish this, it was initially assumed that feedback functioned as reinforcement. Since immediate reinforcement was much more effective in producing acquisition of responses than was delayed reinforcement, immediate feedback was considered to be an integral part of any good instructional strategy. PSI researchers have devoted relatively little time to examining this assumption. Calhoun (1976) compared student performance under delayed and immediate feedback conditions and found that immediate feedback was superior. Unfortunately, Calhoun's study did not examine long-term retention, which is the only measure that has been found to vary consistently as a function of feedback.

Others (Farmer, Lachter, Blaustein, & Cole, 1972; Johnson & Sulzer-Azaroff, 1975) reported findings concerning delayed and immediate feedback in PSI, but their feedback conditions were confounded by method of delivery (proctor-delivered versus written feedback), and so no conclusions regarding the timing of feedback can be drawn from their data.

Recent work by Robin (1978) attempted to examine the effects of differing delays of feedback in a PSI course using essay test items. While there were no differences in acquisition as a function of the delay, students in this study expressed strong preferences for immediate feedback. The author concluded that PSI courses should arrange to provide immediate feedback whenever it is feasible. Unfortunately, Robin did not measure retention as a function of delayed and immediate feedback, and research design used (counter-balanced, within-subject reversal) precludes examination of this aspect. Since previous studies have used primarily multiple-choice items and have shown differences only in retention, no conclusions can be drawn concerning the presence or absence of the delay-retention effect with essay test items. Because the provision of immediate feedback is so costly in terms of student time, proctor time, computer programming, or materials preparation, Robin's recommendation that ". . .it [immediate feedback] should remain an element of most instructional programs" (p. 87) seems unwarranted at this time.

Experimental Conditions

The three experiments described in this report were all conducted under the same general set of experimental conditions. The research occurred in regularly scheduled college courses. Course material was arranged and presented according to the basic principles of the PSI. These principles include (1) frequent repeatable quizzing over small units of material until a mastery criterion is attained, (2) modified self-pacing, and (3) the provision of proctors (tutors) to administer and grade quizzes and provide feedback.

EXPERIMENT I

Experiment I was designed to assess the effects of immediate and delayed feedback on performance in a course using short-answer essay test items.

Approach

Experimental Design and Subjects

Thirty-four students in an introductory cultural anthropology class at San Diego State University were randomly assigned to two feedback groups. The immediate feedback (IMFB) group (N = 18) received feedback 20 minutes after completing the quiz. The delayed feedback (DLFB) group (N = 16) received feedback 48 hours after completing the quiz. The two feedback interval conditions constituted the independent variable. The dependent variables were:

1. Student learning, as measured by performance on first attempts at quizzes.

2. Student retention, as measured by performance on review tests and a final exam.

3. Differential effect of feedback on items correct or incorrect initially but correct later.

4. The amount of student study time.

Test Schedule and Materials

All students were required to take a total of 10 unit quizzes, two review tests, and a final exam. Only four of the unit quizzes were used in the experiment, however. Table 1 shows the sequence in which the experimental unit quizzes were presented.

All questions required short essay answers and all answers were scored as completely wrong or completely right.

The review tests contained five questions from each of the two experimental unit quizzes that preceded them. The final exam questions were taken from the two review tests. One question on the final exam was deleted from the analyses because it was invalid.

Students who did not reach criterion on the experimental unit quizzes were permitted to take an alternate form of the quiz. The alternate form contained the five essay questions from the original quiz that were not used on the review test plus five new questions.

Table 1

Experimental Quiz/Test Schedule (Experiment I)

Week of Semester	Units Covered	Type of Test	Test No.	Number of Questions
2	2	Study quiz	Q2	10
3	3	Study quiz	Q3	10
5	2 & 3	Review test	R1	10
7	6	Study quiz	Q6	10
9	7	Study quiz	Q7	10
10	6 & 7	Review test	R2	10
16	1 thru 7	Final exam	--	20[a]

[a]One question was deleted from the analyses because it was found to be invalid. The final exam included questions from non-experimental units, although these data were not included in the statistical analyses.

Experimental Feedback

Feedback consisted of providing the student with a form with an indication of whether each answer was correct or incorrect. The student was referred to the portion of the text from which the item was drawn.

Criterion. The criterion set for mastery of the material was 70 percent. If students scored lower than 70 percent on a quiz, they were required to take up to two alternate forms to reach criterion. If after three attempts students still had not reached criterion, they received no credit for that unit.

Study Time Sheets

Students maintained, and presented to the proctors, records of the time they spent studying for each test.

Proctors

Each proctor was a graduate student who was responsible for 17 students. Proctor groups included students from both feedback groups. Proctors attended lectures and monitored out-of-class testing sessions, and administered both experimental and non-experimental unit quizzes.

Procedure

Quizzes. Students reported to their proctors, handed in their study time sheets, and received a quiz. While students completed the quiz, proctors recorded the students' study time. When students had finished the quiz, they handed it to the proctor and received their study time sheets back.

1. Students in the IMFB group then waited while proctors corrected their quizzes, recorded the scores, and filled out their feedback information. After this, the proctors gave the students the feedback. Students could keep the feedback until the end of the class period, when it had to be returned to the proctor. Students were allowed to take notes on the textbook references for incorrect items since feedback was not allowed to leave the test area. If students had met criterion, proctors recorded that they had completed the unit when the feedback was returned to them.

2. Students in the DLFB group were excused when they handed in their quizzes and were told their feedback would be ready in 48 hours. Proctors corrected the quizzes, recorded the scores, and prepared the feedback. When students returned, 48 hours later (or as soon after as possible), feedback was given as it was to the immediate feedback group.

Remediations. Remedial quizzes were independently arranged as needed. Proctors recorded the number and form of the alternate quiz they administered. The procedure for giving remedial tests was the same as that for the initial quizzes. All remedial testing was done before the review test covering that material.

Review tests. Review tests were given in the same way as quizzes.

Final exam. Students took the final exam in a traditional test-taking situation; no feedback was given. Students were told their scores immediately, regardless of feedback group.

Quizzes not used in the experiment. Testing was the same for experimental and non-experimental study units. Feedback for the nonexperimental quizzes, however, consisted of the students' corrected quizzes. Students returned the tests at the end of the class period. There were no remediations for non-experimental units.

Analyses

Analyses of variance (ANOVA) tests with type of feedback as the independent variable were conducted on students' reported study time, students' scores on initial quizzes, review tests, and the final examination.

Z-tests of significance were conducted for proportions of items that were: (1) correct and incorrect on the quizzes that were correct on the review test, and (2) correct and incorrect on the review test that were correct on the final exam.

Results of Experiment I

Reliability of Scoring the Short Essay Answers

The overall agreement among scorers was 96.6 percent, ranging from 90 to 100 percent.

Group Performance on Initial Attempts on Quizzes

Each proctor's group contained both delayed and immediate feedback students. A preliminary ANOVA on group performance on the initial attempts on quizzes with two between-group variables—delay of feedback and proctor—revealed no systematic difference between the proctors ($F = 1.66$, $df = 1,30$). Consequently, only feedback delay was considered as a between-groups variable in subsequent analyses.

Table 2 contains the group means for the initial attempt on quiz 2 (Q2), the first quiz used in the experiment. The two feedback groups' performance on Q2 was analyzed by an ANOVA with one between-group variable--delay of feedback. No significant effect was found ($F = .042$, $df = 1,32$), indicating that the two feedback groups did not differ at the start of the experiment.

Table 2

Mean Percent Correct On Quiz 2, Review Tests, and Final Exam (Experiment I)

	Mean Percent Correct					
Feedback Group	Q2	R1	R2	Final Exam Items from R1	Final Exam Items from R2	N
IMFB	92	86	84	70	74	18
DLFB	92	89	82	76	84	16

Group Performance on Review Tests

Table 2 also contains the group means for the review tests. The two feedback groups did not differ significantly in their performance on either review test 1 ($F = .361$, $df = 1,32$) or review test 2 ($F = .048$, $df = 1,30$).

Group Performance on Final Exam

The group means on final exam questions from the review tests are contained in Table 2. The two feedback groups did not differ significantly in their performance on the final exam. Final exam questions taken from review test 1 and review test 2 were analyzed separately (the ANOVA results are $F = .79$, $df = 1,32$ and $F = 3.16$, $df = 1,32$, respectively). There were no significant differences for questions from either review test.

Study Time

Table 3 contains the mean total study time for each feedback group.

When the study times for each unit quiz, review test, and the final exam were analyzed in an ANOVA with feedback delay as the between-groups variable and test scores as the within-groups, or repeated, measure, no significant difference was found between the feedback groups in the amount of study time ($F = .206$, $df = 1,214$).

Number of Remediations

Table 3 also contains the mean total number of remediations taken by each feedback group. The two groups did not differ significantly in the average total number of remediations taken ($F = .236$, $df = 1,32$).

Table 3

Group Means for Total Study Time and Number of Remediations (Experiment I)

Feedback Group	Total Study Time	Total Remediation	N
IMFB	19.6	.8	18
DLFB	21.7	.7	16

Proportions of Items that were Incorrect or Correct on a Review Test that were Correct on the Final Exam

Table 4 contains the proportions of items that were incorrect or correct on a review test that were correct on the final exam. The two feedback groups differed only for items from review test 2, where the DLFB group had a higher proportion of items correct initially and correct later than did the IMFB group ($Z = 3.22$, $p < .01$).

Table 4

Proportion of Items Correct and Incorrect on the Review Tests that were Correct on the Final Exam (Experiment I)

	Proportion of Items Incorrect on a Review Test that were Correct on the Final Exam		Proportion of Items Correct on a Review Test that were Still Correct on the Final Exam	
Feedback Group	Review Test 1	Review Test 2	Review Test 1	Review Test 2
IMFB	.27	.46	.75	.81[a]
DLFB	.25	.56	.82	.94[a]

[a]Immediate versus delay comparison significant at $p < .01$ level.

EXPERIMENT II

Experiment II was designed to determine the independent and combined effects of feedback timing and feedback format.

Approach

Experimental Design

The experimental design for Experiment II was a 2 x 2 factorial design. The independent variables were interval before test feedback (immediate or delayed) and feedback format (feedback only or feedback plus students' degree of correctness).

The dependent variables were three measures of student performance:

1. Student learning, measured in terms of performance on quiz and review test items, both multiple-choice and fill-in.

2. Knowledge retention, measured in two ways.

   a. Loss from review tests to final exam on repeated items (both multiple-choice and fill-in).

   b. Test performance on new items on final exam (both multiple-choice and fill-in).

3. Differential effect of feedback on student performance, measured by:

   a. The proportion of multiple-choice and fill-in items that were answered correctly and incorrectly on study quizzes that were correct on the review tests.

   b. Similar proportions for the final exam items.

Subjects

Subjects were 57 undergraduate students enrolled in four sections of an introductory course in research methodology at California State University, Chico. The course was taught by one instructor, with an additional instructor conducting two of the four laboratory sections that accompanied the lecture part of the course.

Students were randomly assigned to one of the four following groups and remained in that group throughout the semester:

1. Immediate feedback (IMFB) group. Students in the IMFB group (N = 15) received feedback within 20 minutes. The feedback form included the original question and the correct answer.

2. Delayed feedback (DLFB) group. Students in the DLFB group (N = 13) received feedback after a 24-hour interval. The feedback was identical to that provided to the IMFB group.

3. Immediate feedback and rightness/wrongness (IMFBR/W) group. Students in the IMFBR/W group (N = 13) received immediate feedback that included the original question, the correct answer, and an indication of whether the students' answer was right or wrong.

4. Delayed feedback and rightness/wrongness (DLFBR/W) group. Students in the DLFBR/W group (N = 16) received delayed feedback, with the feedback identical to that presented to the IMFBR/W group.

Test Schedule

There were six study quizzes, three review tests, and a final exam, all of which counted toward the course grade. Table 5 presents the sequence in which the tests were given during the semester.

Table 5

Schedule (Experiment II)

Week of Semester	Study Block	Type of Test	Test No.	Number of Items
2	1	Study quiz	Q1	20
TBA[a]	1	Study quiz	Q2	20
4	1	Review test	R1	40
6	2	Study quiz	Q3	20
TBA	2	Study quiz	Q4	20
8	2	Review test	R2	40
10	3	Study quiz	Q5	20
TBA	3	Study quiz	Q6	20
14	3	Review test	R3	40
16	-	Final exam	—	90

[a]To be arranged at student's own pace.

Test Materials

The lecture and the laboratory sessions were conducted independently and the PSI testing examined in this experiment covered the lecture material only.

Study quizzes. Study quizzes had 10 multiple-choice and 10 fill-in or short-answer items. Figure 1 shows a sample of each type of item and the two forms of feedback for each item type.

Review tests. Review tests consisted of all 40 items from the two preceding study quizzes. There was no feedback after the review tests.

Final exam. The final exam consisted of 60 previously-used items and 30 new ones. The previously-used items consisted of 10 multiple-choice and 10 fill-in items from each of the three review tests. The new questions consisted of 12 multiple-choice items, six each from material covered in Blocks I and II, and 18 fill-in items, six from each block. There was no feedback for the final exam.

As the effect of the independent variable decreases, the within-group variance

a. decreases
b. increases
c. does not change systematically in either direction

The correct answer is C

A. Multiple-choice item of the type given to the IMFB and DLFB groups.

As the magnitude of the effect of the independent variable increases, between-group variance _____.

The correct answer is: increases

B. Fill-in item of the type given to the IMFB and DLFB groups.

As the effect of the independent variable decreases, the within group-variance

a. decreases
b. increases
c. does not changes systematically in either direction

The correct answer is: C

You were Right Wrong

C. Multiple-choice item of the type given to the IMFBR/W and DLFBR/W groups.

As the magnitude of the effect of the independent variable increases, between-group variance _____.

The correct answer is: increases

You were Right Wrong

D. Fill-in item of the type given to the IMFBR/W and DLFBR/W groups.

Figure 1. Question types and feedback formats for experiments II and III.

Criterion

The criterion for passing a study quiz was set at 90 percent, that is students needed a score of 18 or higher to pass the quiz. If the criterion was not met on the initial study quiz, two alternate forms of each quiz were available for retakes. If students did not reach criterion after the two retakes, they received the highest of their three scores. Students were permitted to retake tests to better their scores, even if they met criterion on the first attempt. There was no criterion set for review tests or the final exam, and there were no alternate forms or retakes permitted on these tests.

Objectives

Students were given reading assignments and specific learning objectives, written as study questions, for each unit. Students were required to answer the study questions on paper before they could attempt the initial test unit. Proctors collected but did not grade, or even read, these answers at the test sessions.

Tutoring

Three proctors, all graduate students in psychology, were also tutors for the students. Tutoring was available throughout the semester but became mandatory about mid-semester for students who had failed to reach 80 percent (a score of 16) on the first

retake of any study quiz. Students had to obtain a tutor's signature to be permitted to take the second retake of the study quiz.

Procedure

At initial study quiz testing sessions, proctors collected the written study question answers from each student and handed out the tests. Students were not permitted to make an initial attempt at a study quiz without handing in written study questions. Students completed the quiz and handed it in to a proctor to grade. At this point, the procedure differed slightly for each group:

1. Students in the IMFB group were given a feedback form and were told to return to their seats and study it at their own pace. Students were not permitted to keep or take notes on the feedback. When they had finished studying the feedback, students returned it to the proctor and received an objectives (study questions) sheet with the total number they had gotten right on the quiz and a notation of the specific objectives on which they had missed questions. Students then left the class.

2. Students in the DLFB group left immediately after turning in the study quiz. They returned one to two days later at which time the procedure for receiving feedback was the same as for the IMFB group.

3. Students in the IMFBR/W group waited while proctors corrected their tests and marked each question of their feedback right or wrong. Then proctors handed students their marked feedback and the procedure became the same as for the IMFB group.

4. Students in the DLFBR/W group left immediately after turning in the study quiz. When they returned for feedback, in one or two days, the procedure was the same as for the IMFBR/W group.

Retakes. Students took an alternate form of a study quiz if they had not reached criterion on the inital quiz. The restrictions on retakes were that they had to be taken:

1. At least one day after feedback was given.

2. At least two days after the last test was given.

3. No later than six days after the initial study quiz.

The procedure for retaking a quiz was the same as for taking the initial quiz, except that instead of handing in written study questions, students handed in the objectives they had been given at the end of their last study quiz. The statistical analyses for the three experiments described in this report do not include scores from retakes.

Review tests. Prior to each review test, students must have taken the appropriate study quizzes and alternate quiz forms necessary to have attained a criterion of 90 percent. Students were not permitted to take a review test if they had received feedback for an initial study quiz less than 24 hours before the scheduled review test. All review exams were taken in the lecture class except for students who had not completed the appropriate study quizzes. There was no feedback after a review test; all students simply took the test, turned it in, and left the room.

Final exam. The final exam was held during final exam week in a conventional test-taking situation.

Analyses

Factorial analyses of variance (ANOVAs) were conducted to determine whether there were any differences between the four feedback groups in performance on the quizzes, review tests, and final exam.

Z-tests of significance were conducted to determine whether there were any differences between the four feedback groups in:

1. The proportion of correct and incorrect quiz items that were correct on the review test.

2. The proportion of correct and incorrect review test items that were correct on the final exam, for multiple-choice and fill-in items separately.

Results of Experiment II

Reliability of Scoring Fill-in Test Items

The overall reliability for scoring the fill-in test items on the study quizzes was 97.8 percent, ranging from 92.4 to 100 percent.

Group Performance on Study Quiz 1

A preliminary ANOVA on group performance on the first study quiz, with two between-groups variables, feedback timing and feedback format, revealed no significant difference in performance among the four feedback groups ($F = .1152$, $df = 1,53$). The groups did not differ, therefore, at the beginning of the semester.

Table 6 contains the group means for the initial study quiz. The group means on the review test and the final exam are found in Table 7.

Table 6

Mean Number of Correct Answers on Study Quiz 1 (Experiment II)

Feedback Group	$\bar{X}$ Number Correct on Study Quiz 1	N
IMFB	13.92	15
DLFB	12.25	14
IMFBR/W	15.08	12
DLFBR/W	14.00	16

Table 7

Mean Number Correct for Items on Review Tests and Final Exam (Experiment II)

Feedback Group	Multiple-choice			Fill-in		
	R1	R2	R3	R1	R2	R3
Mean Numbers of Items Right on Review Tests (R1-R3)						
IMFB	9.00	9.13	8.40	9.40	9.07	8.73
DLFB	8.36	9.07	8.21	9.14	8.79	8.86
IMFBR/W	8.58	9.21	8.31	9.25	9.07	8.23
DLFBR/W	8.68	8.88	9.13	9.38	9.00	9.19
Mean Numbers of Items from Review Tests that Were Right on Final Exam						
IMFB	8.73	8.80	8.00	8.67	8.20	8.67
DLFB	8.36	8.36	7.50	7.86	8.00	8.43
IMFBR/W	8.92	8.21	7.77	8.50	7.93	7.77
DLFBR/W	8.81	8.06	8.00	9.06	8.50	9.13

Group Performance on Multiple-choice Items

Gain from Study quiz to review test. All students scored higher on multiple-choice items on the review tests than they had scored on the same items on the study quizzes. An ANOVA with two between-groups variables--feedback delay and feedback format--was performed for each of the six study quizzes. The repeated measures were the scores on the quizzes and review tests. A significant effect of scores, with review test scores being higher than study quiz scores, was found for all quizzes. (Typical ANOVA results are $F(1,53) = 28.4$, $p < .001$ for study quiz and review test 1.)

The two groups receiving immediate feedback gained significantly more than the two groups receiving delayed feedback ($F(1,53) = 4.46$, $p < .05$) on multiple-choice items compared between the first study quiz and the review test. But feedback delay was not significant for multiple-choice item comparisons with review tests for any other quizzes. When analyzed alone the immediate and delayed groups did not differ systematically in their performance on multiple-choice items.

Feedback format affected student performance on multiple-choice items of study quizzes 5 and 6. The IMFBR/W and DLFBR/W groups scored lower on study quiz 5 but did better on the review test, than the IMFB and DLFB groups ($F(1,54) = 5.23$, $p < .05$).

The mean numbers of multiple-choice items correct on study quiz 5 and review test 3 are given below.

Feedback Groups	Study Quiz 5	Review Test 3
IMFB and DLFB	7.1	7.8
IMFBR/W and DLFBR/W	6.3	8.5

This finding was somewhat negated by the performance on study quiz 6, where the IMFB and DLFB groups did better on both the study quiz and the review test than did the IMFBR/W and DLFBR/W groups ($F(1,54) = 4.15$, $p < .05$).

Loss from review test to final exam. The scores obtained on multiple-choice items used on both the review tests and the final exam were analyzed using an ANOVA with two between-groups variables (feedback timing and feedback format). Scores were higher on the review test than on the final exam for items from review tests 2 and 3 (for 2, $F(1,55) = 12.47$; for 3, $F(1,54) = 11.09$, both significant at $p < .001$). There was no systematic relationship between the loss from any review test to the final exam and the timing or format of the feedback.

New multiple-choice items of final exam. The final exam scores obtained on new multiple-choice items covering material from the first eight weeks of the course were analyzed using an ANOVA. The between-groups variables were feedback timing and feedback format. The within-group measure was the score on new items. There was no difference among the four feedback groups.

Group Performance on Fill-in Items

Gain from Study quiz to review test. The results for fill-in items were similar to those for multiple-choice items. An ANOVA was performed for each of the six study quizzes using feedback format and feedback timing as between-groups variables. Scores on quizzes and review tests were used as the within-group variable. Scores on the fill-in items were significantly higher on the review tests than they were on the study quizzes.

Performance on fill-in items used on quiz 5 and on the review test differed significantly for the two feedback groups ($F(1,54) = 11.27$, $p < .001$). The IMFB and IMFBR/W groups scored higher on the quiz, but gained less on the review test, than did the DLFB and DLFBR/W groups. The mean numbers of items correct on quiz 5 and review test 3 are as follows:

Feedback Groups	Study Quiz 5	Review Test 3
IMFB and IMFBR/W	7.3	7.8
DLFB and DLFBR/W	6.2	8.9

Loss from review test to final exam. The scores obtained on fill-in items used on both the review tests and on the final exam were analyzed using an ANOVA. The between-groups variables were feedback timing and format. A significant loss was noted from review tests 1 and 2 to the final exam (for 1, $F(1,53) = 19.81$, $p < .001$; for 2, $F(1,55) = 29.75$, $p < .001$), but there was no loss from review test 3. These losses did not vary for the different feedback intervals or formats.

New fill-in items on the final exam. The performance of the four feedback groups did not differ systematically on new fill-in items on the final exam.

Proportions of Items that were Incorrect or Correct on a Study Quiz that were Correct on the Next Review Test

The effect of feedback delay and feedback format on the proportion of multiple-choice and fill-in items that were correct on the review test was analyzed separately for items that were correct and that were incorrect on the study quizzes. As in Experiment I, each student's response on the study quizzes were divided into those that were correct and those that were incorrect, and the proportions of each of these that were correct on the review tests were computed separately for the four feedback groups. Comparisons were made between the four feedback groups and for each item type, multiple-choice and fill-in. Tables 8 and 9 contain these proportions. No systematic effects were found.

Table 8

Proportions of Multiple-choice Items Incorrect or Correct on a Study Quiz that Were Correct on the Next Review Test (Experiment II)

Feedback Group	Proportion of the Items that Were Wrong on Quizzes (Q1-Q6) that Were Right on the Next Review Test						Proportion of the Items that Were Right on Quizzes (Q1-Q6) that Were Still Right on the Next Review Test					
	Q1	Q2	Q3	Q4	Q5	Q6	Q1	Q2	Q3	Q4	Q5	Q6
IMFB	.76	.55	.56	.94	.47	.92	.96	.96[a]	.93	.94	.88	.96
DLFB	.74	.57	.70	.77	.52	.92	.92	.87[a]	.94	.97	.87	.96
IMFBR/W	.75	.67	.65	.75	.59[b]	.81	.98	.86	.95	.97	.89	.95
DLFBR/W	.79	.70	.84	.83	.85[b]	.95	.94	.82	.91	.92	.92	.96

[a]The difference between the performance of the IMFB and DLFB groups on study quiz 2 was significant at $p < .05$.

[b]The difference between the performance of the IMFBR/W and DLFBR/W groups on study quiz 5 was significant at $p < .05$.

Table 9

Proportions of Fill-in Items Incorrect or Correct on a Study Quiz that Were Correct on the Next Review Test (Experiment II)

Feedback Group	Proportion of the Items that Were Wrong on Quizzes (Q1-Q6) that Were Right on the Next Review Test						Proportion of the Items that Were Right on Quizzes (Q1-Q6) that Were Still Right on the Next Review Test					
	Q1	Q2	Q3	Q4	Q5	Q6	Q1	Q2	Q3	Q4	Q5	Q6
IMFB	.72	.65	.73	.76	.66	.82	.92	.91[a]	.95	.94	.90	.98
DLFB	.74	.71	.70	.73	.73	.94	.84	1.00[a]	.94	.91	.94	.96
IMFBR/W	.77	.70	.73	.80	.56[b]	.76	.93	.88	.96	.97	.86[b]	.93
DLFBR/W	.88	.64	.70	.75	.76[b]	.89	.88	.92	.97	.95	.97[b]	.96

[a]The difference between the performance of the IMFB and DLFB groups on study quiz 2 was significant at $p < .05$.

[b]The difference between the performance of the IMFBR/W and DLFBR/W groups on study quiz 5 was significant at $p < .05$.

The IMFB group had a higher proportion of multiple-choice items from quiz 2 correct on the study quiz and correct on the review test than did the DLFB group ($Z = -2.2519$, $p < .05$), as shown in Table 8.

For fill-in items from the same study quiz, the results were the opposite (Table 9). The DLFB group had a higher proportion of fill-in items that were correct on quiz 2 and still correct on the review test than did the IMFB group ($Z = 2.9632$, $p < .01$).

The IMFB and DLFB groups did not differ in the proportion of multiple-choice and fill-in items that were wrong initially and right later.

There was no difference between the IMFBR/W and DLFBR/W groups in the proportion of multiple-choice items right on the study quiz and right later on the review test.

Delayed feedback seemed to enhance the performance of the DLFBR/W group on items that were wrong initially. For study quiz 5, the DLFBR/W group had a higher proportion of items that were wrong initially but right later (multiple-choice, $Z = -2.9130$, $p < .01$; fill-in, $Z = -1.9636$, $p < .01$).

Proportions of Items that were Incorrect or Correct on a Review Test that were Correct on the Final Exam

As shown in Table 10, neither feedback timing nor feedback format had any significant effect on final exam scores.

Table 10

Proportions of Items Incorrect or Correct on a Review Test that Were Correct on the Final Exam (Experiment II)

Feedback Group	Proportion of Items that Were Wrong on the Review Tests (R1-R3) that Were Right on the Final Exam						Proportion of Items that Were Right on the Review Tests (R1-R3) that Were Still Right on the Final Exam					
	Multiple-choice			Fill-in			Multiple-choice			Fill-in		
	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3
IMFB	.56	.62	.46	.56	.29	.37	.91	.91	.87	.89	.88	.94
DLFB	.50	.64	.24	.45	.35	.38	.91	.85	.86	.81	.86	.90
IMFBR/W	.59	.36	.18	.44	.23	.25	.94	.86	.90	.88	.85	.90
DLFBR/W	.62	.58	.36	.40	.25	.38	.92	.84	.84	.94	.92	.96

EXPERIMENT III

Experiment III was also designed to measure the effects of timing of feedback, but not feedback format. Three other modifications from Experiment II were:

1. The test schedule for experiment III was designed to eliminate review sessions before exams while at the same time allowing students more time in class to take tests and receive feedback. The number of tests was still limited to an initial test and two alternate forms for retakes.

2. Criterion for study quizzes was lowered from 90 to 80 percent so that students could progress faster through the testing schedule. It was unrealistic to expect all students to reach a 90 percent criterion with this subject matter.

3. The testing schedule was changed to give more time between even numbered quizzes and review tests for students who were unable to meet out-of-class testing sessions. The changed testing also allowed those who were able to do so to move more quickly through the testing program by elimination of the review sessions before exams.

Approach

Experimental Design and Subjects

The experimental design for experiment III was a two group design with the independent variable being interval before feedback, either immediate or delayed. The dependent variables were the same measures of learning used in experiment II.

The subjects, 30 undergraduate students enrolled in two sections of an introductory course in research methodology, were randomly assigned to one of the two following groups and remained in that group throughout the semester.

1. Immediate feedback (IMFB) group. Students in the IMFB group (N = 16) received feedback after a 20 minute interval.

2. Delayed feedback (DLFB) group. Students in the DLFB group (N = 14) received feedback after a 24-hour delay.

Test Schedule

The test schedule is given in Table 11. Students in experiment III could take a new quiz without having reached criterion on the previous one.

Table 11

Test Schedule (Experiment III)

Week of Semester	Unit(s) Covered	Type of Test	Test No.	Number of Items
2	1	Study quiz	Q1	20
TBA[a]	2	Study quiz	Q2	20
4	3	Study quiz	Q3	20
5	1 & 2	Review test	R1	40
TBA	4	Study quiz	Q4	20
8	5	Study quiz	Q5	20
9	3 & 4	Review test	R2	40
TBA	6	Study quiz	Q6	20
14	5 & 6	Review test	R3	40
16	All	Final exam	—	90

[a]To be arranged at student's own pace.

Testing Materials

Study quizzes. Study quizzes were the same as in experiment II.

Feedback. Feedback, provided by feedback forms, was similar to that given to the IMFB and DLFB groups in experiment II; for multiple-choice items the letter of the correct alternative was given; for fill-in items, the correct short answer was given.

Review exams. Review exams were similar to those used in experiment II.

Final exam. The final exam was similar to the one used in experiment II, although different items were used. Four multiple-choice items were randomly discarded to ensure an equal number of multiple-choice and fill-in questions.

Criterion. Students were required to have taken the two scheduled quizzes before they could take the corresponding review test, whether or not the study criterion of 80 percent had been met when the review test was scheduled.

Objectives. Learning objectives were the same as in experiment II.

Tutoring

Tutors in experiment III were five undergraduate students who had completed the course the previous semester. They were available for tutoring throughout the semester and they also proctored the out-of-class test-taking. It was suggested that students see a tutor if they obtained 60 percent or less (12 items or less) correct on a scheduled quiz. Students were required to see a tutor and obtain a tutor's signature if they scored less than 60 percent on the first retake. Students had to present the tutor's signature to take the second retake of the test.

Procedure

As with experiment II, lectures covered material in the objectives but no lecture sessions were devoted to reviewing the material before the review tests.

Study quizzes. Study quizzes 1, 3, and 5 were given in lecture. Quizzes 2, 4, and 6 could be taken during the last half of lecture, during lab periods, or at specified hours outside of class. The testing procedure was the same as it was for experiment II.

Students in the IMFB group were given feedback forms to study while proctors corrected their tests. Students studied feedback at their own pace, spending as much time as they wanted. They were not permitted to take notes on it or keep the sheets. Students then returned the feedback to the proctor and received an objectives sheet (for the unit they were tested on) with their number correct on it and notations indicating any areas they had missed.

Students in the DLFB group left after turning in the quiz. They returned one to two days later for feedback. The procedure at that time was the same as for the immediate feedback group.

Retakes. If students did not meet the criterion the the first time they took a quiz they had to take an alternate form of the failed quiz. Retakes had the following constraints:

1. Retakes could be taken no sooner than two days after the failed quiz had been taken.

2. Two hours after feedback is received.

3. Retakes had to be taken no later than six days after the failed quiz.

When students retook the quizzes, they gave the proctors the objectives they had been given after taking their last quiz so the proctor could verify that the test rules were being observed. The procedure for taking an alternate form of a quiz was the same as for taking a scheduled quiz.

Review tests and final examinations. Procedures administering the review tests and final examinations were identical to those in experiment II.

Analyses

ANOVAs conducted for experiment III were similar to those done for experiment II, but with only the two feedback groups instead of four.

Results of Experiment III

The overall reliability for scoring the fill-in items was 96.5 percent, ranging from 90.3 to 100 percent.

Group Performance on Quiz 1

The two feedback groups did not differ at the beginning of the semester. A preliminary ANOVA on group performance on quiz 1, using feedback timing as the between-groups variable, revealed no significant difference between the two groups ($F = .0141$, $df = 1,28$). The mean numbers correct for quiz 1 were 14.6 for the IMFB group and 14.8 for the DLFB group.

The group means for each item type on the review test and the final exam are found in Table 12.

Table 12

Mean Numbers Correct for Items on Review Tests and Final Exam (Experiment III)

Feedback Group	Multiple-choice			Fill-in		
	R1	R2	R3	R1	R2	R3
Mean Numbers of Items Right on Review Tests (R1-R3)						
IMFB	7.00	7.82	7.08	6.38	8.09	7.33
DLFB	7.42	8.21	6.89	6.17	7.92	6.89
Mean Numbers of Items from Review Tests that Were Right on Final Exam						
IMFB	6.77	8.00	5.92	5.23	6.82	6.58
DLFB	7.00	7.15	6.11	5.33	6.85	6.89

[a]Maximum score was eight.

Group Performance on Multiple-choice Items on Study Quizzes and Review Test

An ANOVA was performed using feedback timing as the between-groups variable. The within-group variables were quiz and review test scores and quiz 1 vs. quiz 2 scores. The ANOVA was performed for multiple-choice items from each of the three review tests. For multiple-choice items from quizzes 1 and 2, all students got higher scores on

the review test than they did on the quizzes ($F(1,28) = 8.68$, $p < .01$). The two feedback groups did not differ in their performance on any of these measures.

Items from quiz 1 were right more frequently than were items from quiz 2 ($F(1,28) = 30.65$, $p > .01$).

There were no significant feedback effects found for multiple-choice items from quizzes 3 and 4. The two feedback groups did not differ in their performance on multiple-choice questions from any of the quizzes and the review test scores were not significantly higher than the study quiz scores for review tests 2 and 3.

Items from quiz 6 were more frequently correct than were items from quiz 5 ($F(1,20) = 9.23$, $p < .01$).

Group Performance on Fill-in Items on Quizzes and Review Tests

The fill-in items were analyzed using the same analysis as was used for multiple-choice items. The analysis resulted in even fewer differences. For study quizzes 1 through 4, no significant effects were found for feedback delay, tests, questions, or for any interaction between these. On study quizzes 5 and 6, performance on the review test was significantly better than on the study quizzes ($F(1,20) = 6.45$, $p < .01$). There was an interaction between test and questions, with items from quiz 5 being answered correctly on review test 3 slightly more often than on the quiz. Items from quiz 6 were answered correctly on review test 3 much more often than on the quiz ($F(1,20) = 4.88$, $p < .05$).

The test-by-question interaction can be seen from the mean scores for items that were used on study quiz 5 or 6 and again on review test 3, as follows:

Quiz Items	Score on Quiz	Score on Review Test 3
From quiz 5	6.7	6.8
From quiz 6	6.0	7.8

Loss from Review Test to Final Exam for Multiple-choice and Fill-in Items

The performance on multiple-choice and fill-in items on both the review test and the final exam was analyzed using an ANOVA with one between-groups variable, feedback timing. The within-group variables were tests (review tests and final exam) and questions (multiple-choice or fill-in). Scores were higher on the review tests than they were on the final exam for all review test questions, and multiple-choice questions were correct more frequently than fill-in questions on material from review test 1 ($F(1,23) = 44.67$, $p < .01$). There were no systematic differences in the performance of the two feedback groups.

Number of Remediations

A simple between-groups ANOVA was performed for the number of remediations for each quiz (Table 13). There were no significant differences between the two feedback groups.

Table 13

Mean Numbers of Remediations
(Experiment III)

Feedback Group	Mean Numbers of Remediation Tests Required to Reach Criterion after each of the Scheduled Study Quizzes (Q1-Q6)					
	Q1	Q2	Q3	Q4	Q5	Q6
IMFB	1.00	.81	.75	.42	.92	.85
DLFB	.64	.79	.71	.50	1.00	.56

Proportions of Items that were Incorrect or Correct on a Study Quiz that were Correct on the Next Review Test

As in experiments I and II, the effect of feedback-type on the proportions of items that were incorrect or correct on the quizzes that were later correct on a review test was analyzed (Tables 14 and 15), When z-tests of proportions were performed, there were no significant differences for the two feedback groups on any comparisons.

Table 14

Proportions of Multiple-choice Items that Were Incorrect or Correct on the Study Quizzes that Were Correct on the Next Review Test (Experiment III)

Feedback Group	Proportion of the Items that Were Wrong on Quizzes (Q1-Q6) that Were Right on the Review Test						Proportion of the Items that Were Right on Quizzes (Q1-Q6) that Were Still Right on the Review Test					
	Q1	Q2	Q3	Q4	Q5	Q6	Q1	Q2	Q3	Q4	Q5	Q6
IMFB	.75	.47	.62	.52	.43	.65	.93	.87	.87	.89	.80	.82
DLFB	.58	.51	.66	.74	.58	.57	.92	.90	.85	.89	.84	.88

Table 15

Proportions of Fill-in Items that Were Incorrect or Correct on the Study Quizzes that Were Correct on the Next Review Test (Experiment III)

Feedback Group	Proportion of the Items that Were Wrong on Quizzes (Q1-Q6) that Were Right on the Next Review Test						Proportion of the Items that Were Right on Quizzes (Q1-Q6) that Were Still Right on the Next Review Test					
	Q1	Q2	Q3	Q4	Q5	Q6	Q1	Q2	Q3	Q4	Q5	Q6
IMFB	.41	.58	.52	.32	.40	.62	.86	.80	.91	.85	.88	.92
DLFB	.55	.63	.53	.55	.38	.43	.77	.80	.90	.90	.83	.92

Proportions of Items that were Incorrect or Correct on Review Tests that were Correct on the Final Exam

Comparisons were made between the IMFB and DLFB groups for the proportions of items that were incorrect or correct on a review test that were correct on the final exam (Table 16). Z-tests revealed no significant differences between the two feedback types.

Table 16

Proportions of Items Incorrect or Correct on Review Tests that Were Correct on the Final Exam (Experiment III)

Feedback Group	Proportion of Items that Were Wrong on the Review Tests (R1-R3) that Were Right on the Final Exam						Proportion of Items that Were Right on the Review Tests (R1-R3) that Were Still Right on the Final Exam					
	Multiple-choice			Fill-in			Multiple-choice			Fill-in		
	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3
IMFB	.38	.46	.26	.29	.19	.31	.91	.90	.73	.77	.81	.78
DLFB	.57	.27	.31	.41	.30	.37	.90	.81	.79	.74	.79	.85

CONCLUSIONS

This series of experiments revealed no pattern of significant differences in long-term knowledge retention as a function of feedback interval, test-item type, or quality of feedback. These findings extend the work of Calhoun (1976), Farmer, et al. (1972), Johnson and Sulzer-Azaroff (1975), and Robin (1978), but do not support the conclusion that performance under immediate feedback is superior, or that PSI (or other similar instructional systems) should provide immediate feedback whenever possible.

The experiments likewise fail to support the findings of superiority of delayed feedback that have been obtained by many other researchers in experimental and more conventional classroom settings (c.f., Sturges, 1969, 1972, 1978; Surber & Anderson, 1975). Classroom procedures in PSI differ considerably from those in conventional classrooms, and it is reasonable to look at the differences between the PSI and conventional procedures as a source of this discrepancy in findings..

Experiments II and III examined two possible causes of the discrepancy between the findings of experiment I and the results of earlier research: test-item type and quality of feedback. Most PSI courses have used essay tests, and these may prompt students to adopt study habits or test-taking strategies that differ from the ones they employ with multiple-choice or fill-in tests. These differences might make delayed feedback less effective in essay tests; or perhaps students must attend more carefully to feedback (delayed or immediate) to determine the correctness of their answers because of the length and complexity of essay items. This does not appear to be the case, however, as there were no differences in the performances of the IMFB and DLFB groups on the essay items used in experiment I.

Similarly, feedback in a typical PSI course usually consists of an indication of the correctness of the response and information concerning the source of the test-item so the student may refer to the text for the correct answer. Most research showing the superiority of delayed feedback, however, has included the correct answer in the feedback. Perhaps the effectiveness of feedback in PSI is so reduced by omitting the correct answers that it does not matter whether it is delayed or immediate. This, too, does not appear to be the case as informational quality of feedback, as varied in these experiments, produced no pattern of differences.

Two other possibilities exist. (1) Proctors administer feedback in PSI but not in conventional classrooms. Perhaps the proctor directs the student's attention more carefully to the feedback, thus obviating the differences due to delay interval. This possibility is plausible if Sturges (1972) is correct that the superiority of delayed feedback is due to the fact that students typically study delayed feedback more closely. (2) PSI includes repeated testing to mastery on quiz units, and conventional testing typically does not. It may be that repeated exposure to the material and to the feedback makes the timing of feedback a less potent variable.

These present experiments shed no light on these latter two possibilities. Further research is necessary to clarify these issues.

RECOMMENDATION

Despite the lack of a definitive, theoretically satisfying explanation for the findings obtained in this series of experiment, it is still possible to make recommendations concerning the timing of feedback in instructional settings. Because the findings across these experiments were consistent, they provide no evidence of superiority of either delayed or immediate feedback in producing immediate knowledge acquisition or long-term retention. The use of immediate feedback in Navy training is not warranted, therefore, when cost and convenience of administration are important considerations.

REFERENCES

Calhoun, J. F. The combination of elements in the personalized system of instruction. Teaching of Psychology, 1976, 3, 73-76.

Farmer, J., Lachter, G. D., Blaustein, J. J., & Cole, B. K. The role of proctoring in personalized instruction. Journal of Applied Behavior Analysis, 1972, 5, 401-405.

Johnson, K. R., & Sulzer-Azaroff, B. The effects of different proctoring systems upon student examination performance and preference. In Research and Technology in Colleges and University Teaching. J. Johnson, & G. W. O'Neill (Eds.). Gainesville, FL: University of Florida, Society for Behavioral Technology and Engineering, 1975, 159-185.

Keller, F. S. "Goodbye, teacher. . ." Journal of Applied Behavior Analysis, 1968, 1, 79-89.

Kulhavy, R. W., & Anderson, R. C. Delay-retention effect with multiple-choice tests. Journal of Educational Psychology, 1972, 63, 505-512.

Moore, A. J. Delay of feedback and the acquisition and retention of verbal materials in the classroom. Journal of Educational Psychology, 1969, 60, 339-342.

Robin, A. L. The timing of feedback in personalized instruction. Journal of Personalized Instruction, 1978, 3, 81-88.

Sassenrath, J. M., & Yonge, G. D. Delayed information feedback, feedback cues, retention set, and delayed retention. Journal of Educational Psychology, 1968, 59, 69-73.

Sassenrath, J. M., & Yonge, G. D. Effects of delayed information feedback and feedback cues in learning and retention. Journal of Educational Psychology, 1969, 60, 174-177.

Sturges, P. T. Verbal retention as a function of the informativeness and delay of informative feedback. Journal of Educational Psychology, 1969, 60, 11-14.

Sturges, P. T. Information delay and retention: Effect of information in feedback and tests. Journal of Educational Psychology, 1972, 63, 32-43.

Sturges, P. T. Delay of informative feedback in computer-assisted testing. Journal of Educational Psychology, 1978, 70, 378-387.

Surber, J. R., & Anderson, R. C. Delay-retention effect in natural classroom settings. Journal of Educational Psychology, 1975, 67, 170-173.

Van Matre, N. H. Computer-managed instruction in the Navy: I. Research background and status (NPRDC Spec. Rep. 80-33). San Diego: Navy Personnel Research and Development Center, September 1980.